Deception in Program Evaluation Design

31 October, 2014

Scott Cheney-Peters

Introduction

As a Russian regiment, exhausted from a 20-mile march through the Austrian countryside, reaches its comrades and rest, it receives word that it must prepare for a sunrise inspection by the alliance's commander-in-chief. The soldiers—believing that they are to look their best in the morning—grudgingly spend the night mending and cleaning their parade uniforms. Unbeknownst to them, the actual intent of the inspection is to convey to their Austrian allies their worn-down state and inability to immediately join battle. At the last minute, the soldiers are told they must cast aside their freshly polished outfits and re-don their tattered greatcoats and dirty marching gear. The inspection thus conveys the desired message about the "sorry condition" of the troops, and they are given a chance to rest before returning to action.[1]

The above account, from Tolstoy's *War and Peace*, highlights the interplay of deception and program assessments, and the importance of considering the possible opportunities for deceit when establishing standards and criteria for evaluation, and the evaluation design itself. In this particular instance, the "program" is the unit's readiness and the program manager (the Russian commander-in-chief) is able to use the design of the program assessment (the inspection) to manipulate its outcome to fool his Austrian allies. This is because he knows that in the absence of any other intelligence, the assessment will rely on the commonly accepted standards for

---

Developed from coursework for Center for Development of Security Excellence, 509 – Program Assessment and Evaluation

[1] Tolstoy, War and Peace, 112-115.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
31 OCT 2014	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Deception in Program Evaluation Design	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
Scott Cheney-Peters	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Center for Development of Security Excellence, Defense Security Service, 938 Elkridge Landing Road Linthicum, MD 21090	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

From the stages of criteria and standards selection onward to evaluation design, program managers have a range of options to deceptively influence the outcome of assessments. Other stakeholders, and those wishing to mitigate and minimize manipulation, must remain on guard for its possibility and take proactive steps to reduce the possibility of deceit. These range from the use of open and transparent feedback to ensuring the independence of assessors to red cells identifying possible vulnerabilities. As long as the stakes in a program assessment may influence decisions or influence perceptions, there is every reason to believe that some level of deception will continue in program reporting. Even when manipulation is unintentional, perhaps the result of unconscious prejudgment or preference, the effects on an assessmentâs outcome can be similar. Luckily, stakeholders interested in assessments as a true reflection of a programâs state have a variety of methods at hand to mitigate their impacts. Even in assessments devoid of conscious deceit, the lessons drawn can help improve the fidelity and reliability of the evaluationâs results. Yet as with much of the field, a lot of the recommendations are easier said than done.

15. SUBJECT TERMS

Program Evaluation, Program Design, Program Assessment, Program Management, Evaluation Design, Assessment Design, Deception

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**17**	

determining whether the troops are ready to fight—the cleanliness of whatever uniform they happen to be wearing.[2]

Program managers, and those further up a program's accountability chain, today face many of the same pressures regarding evaluations as they did in the Napoleonic era. Yet despite the marked professionalization of the field of program assessment, program managers and their superiors maintain an ability to deceive evaluators as to the true state of their programs by means of selecting criteria and standards against which to judge programs, as well as the way program evaluations are designed. In fact, others, such as evaluation consultant Michael Patton, believe that the increased role of evaluations as a management and corrective tool means they have "also become more subject to manipulation and abuse."[3]

This paper will examine the causes of program evaluation manipulation and the ways in which it might occur. This will help us draw broader lessons for establishing assessment standards, criteria, and design. Even when manipulation is unintentional (perhaps the result of unconscious prejudgment or preference) the effects on an assessment's outcome can be the same. Therefore the recommendations developed can also help improve the fidelity and reliability of evaluations devoid of conscious deceit.

Motivation to Manipulate

What would cause an individual or organization to attempt to disguise the true state of a program? To understand this it is first necessary to appreciate the purposes of program assessment. One school of thought contends that a main objective is to "influence decisions"—

[2] One could argue that the marching uniforms in fact told the more truthful tale of the program's conditions than the clean parade uniforms would have done.
[3] Patton, *Utilization-Focused Evaluation*, 26. Cites attempts to manipulate the reception and understanding of findings on climate research and intelligence reports.

whether determining the future of the program, resource allocation, or subsequent choices otherwise impacting stakeholders.[4]

In such a construct a stakeholder might be driven to manipulate a program's assessment under the belief that the assessment could, for example, directly affect an individual's employment or salary, or that a negative assessment might spur a decision to either boost or cut resources depending on the context and regulations, or to cut the program completely. In short, a stakeholder might attempt to manipulate the program assessment under the belief the assessment could push a decision in a more favorable direction.

In the case of programs with intended external beneficiaries, such as government aid programs, the beneficiaries typically have different outlooks and different motivations for deception than those who manage the programs. These motivations can nonetheless be illustrative of how deception can skew assessments**.** In an article in the *Journal of the European Economic Association,* Martinelli and Parker looked at a poverty reduction program and uncovered widespread "under-reporting of goods and desirable home characteristics"[5] and, unsurprisingly, tied this directly to mis-reporters' understanding of the benefits they would receive if their income was determined to be under a certain threshold. As another example, in an a 1992 report, GAO's inspector general found that roughly 21% of all tenants in a low-income housing program were guilty of underreporting their income to authorities determining eligibility.[6] As with participants, stakeholders can also be driven to deception when under the belief a decision rests on the outcome of the assessment of the program as a whole or a particular program element.

[4] See for example Habict, Victoria, and Vaughn, "Evaluation Designs for Adequacy, Plausibility, and Probability of Public Health Programme Performance and Impact," 10-18
[5] Martinelli and Parker, "Deception and Misreporting in a Social Program," 886-908.
[6] U.S. General Accounting Office. GAO/HRD-92-60.

Another motivation to deceive derives from another objective of program assessments, often characterized as an attempt to influence perception or "communication" about the state of something. For example, the outcome of an assessment could have a reputational impact for the program or the program's manager and sponsor agency. While this could, as in the above scenario from Tolstoy, very well also influence decisions, those decisions are not necessarily the purpose of the assessment nor the motivation of the stakeholder to manipulate its outcome.

In addition to their examinations of under-reporting in a poverty program, Martinelli and Parker also found over-reporting of goods linked to social status—even at the cost of potentially losing out on program benefits.[7] Martinelli and Parker draw the lesson that an "embarrassment motive," in this case embarrassment at lacking things signifying social status, can spur deception. While focused on a program's participants, this finding can be applied to managers and demonstrates how the motivation to influence perception is distinct from and can in fact negate resource-maximizing attempts to influence decisions.[8]

Yet another example is helpful. The GAO report "Ballistic Missile Defense: Records Indicate Deception Program Did Not Affect 1984 Test Results" details a related scenario in which program evaluation deception is aimed at the perceptions of a competitor, in this case the Soviet Union. The report discusses a series of Army missile interceptor tests designed so that in the case of an interceptor near-miss the target would explode anyway and fool the Soviet sensors expected to monitor the test. According to the GAO, the "deception was seen as a means of impacting arms control negotiations and influencing Soviet spending."[9]

This example is doubly insightful in that the GAO was asked to investigate due to concerns that the intentional deception of the Russians also served to unintentionally deceive

[7] Martinelli and Parker, "Deception and Misreporting in a Social Program," 886-908.
[8] Ibid.
[9] U.S. General Accounting Office. GAO/NSIAD-94-219.

program stakeholders in Congress during a later apparently successful test interception, and therefore provided a deceptive foundation for the decision to further fund the interception program. While the GAO found the claims that Congress was deliberately deceived to be unsubstantiated, they illustrate the possibility of misdirected deception, in addition to intentional and unintentional deception.

In all of the above cases, when deliberately undertaken, the motivation to deceive lies in what the stakeholder expects to be the effect of the outcome of the assessment, whether an impact on a decision or a perception. In many instances, it's possible these motivations overlap.

Designing to Deceive

Given the sometimes compelling motivations for stakeholders to "game the system" for the chance to achieve a preferred outcome, how would they go about doing so? From the development of programmatic standards to the criteria selection for evaluations to their design, there is a multitude of points across program evaluation which might be targeted for manipulation. The first are establishing and selecting the standards to best portray the state of a program, its efficacy, or efficiency.

Forward-thinking manipulators might have the opportunity to try to influence what is measured (the criteria) and the measurements themselves, well before program evaluation design. To maximize the utility of an assessment it is vital to involve knowledgeable stakeholders in the selection of criteria and standards, both to ensure the results are relevant to decision-makers' and also because they are a source of insight on the best items to evaluate and standards by which to judge them. As Havens states, "program evaluation serves little purpose if

it exists in a world unto itself, isolated from the process of program management."[10] But inclusion also creates several openings for stakeholders to attempt to steer criteria and standards towards those that will influence the outcome and impacts of a future program evaluation in the manner they choose, and away from those that do the opposite.[11]

As noted in the U.S. GAO's *Designing Evaluations* handbook, all responsible evaluation designers have to make trade-offs between the sophistication of an assessment and its expected costs in time, money, and other resources.[12] Designers must constantly ask whether the value of an expected increase in fidelity and insights is worth additional costs when creating assessments, whether the present expected results are good enough, and whether there are ways to make the assessment cheaper yet still effective. Motivated stakeholders can use this inherent focus on cost-consciousness in program evaluation to their advantage. For example, they might seek to increase the real or perceived cost of using a specific standard during criteria selection, and conversely to argue that those items that are to them desirable for inclusion will be cheap and easy to measure.

A related approach would be to establish standards so low or high that the vast majority of programs pass or fail, thereby helping to disguise the differences in effectiveness or efficacy among them. This is possible in a scenario in which the decision-maker to whom the program manager is accountable sees or cares only about pass/fail criteria. Such a focus on a single threshold might be driven cost-considerations, especially if it is synonymous with a sole metric, but the motivated stakeholder could also advocate such criteria selection.

Next, stakeholders could insist on measurements requiring a high level of expertise. This could either help to drive up the costs of evaluating an undesirable set of criteria, or it could

[10] Havens, "Program Evaluation and Program Management," 480-485.
[11] As Havens describes it, "…a desire to keep the evaluators out of mischievous activities."
[12] U.S. Government Accountability Office. GAO-12-208G.

necessitate the evaluator possess skill set limited to a small number of personnel whom the stakeholders can count upon to protect their interests. This also aligns with a related approach: attempting to use personal relations to aid in manipulation. Examples might be the establishment of preferred standards and criteria or outright bias in subjective judgments.[13]

One of the areas most vulnerable to such personal bias manipulation is the evaluation of complex performances, as they tend to rely on qualitative rather than purely data-driven analysis. Mislevy defines a complex performance as the interaction of a person and situations of various kinds, for example, "making sense of a mass of disparate material in an art portfolio."[14] In such a situation, where an evaluator will ultimately try to determine the effects and influences of a program on the outcome or behavior of observed complex situations, a stakeholder could try to ensure the criteria is based in large part on the subjective judgment of the assessor. A program manager undertaking this manipulation must be certain he or she will draw a predictably favorable assessor, however, or run the risk of the gamble failing and being assessed worse than in an objective evaluation.

As mentioned, standards and criteria selection are not the only routes for deception to take hold. The design of a program assessment also offers fertile ground. Akin to the move to rely on qualitative judgments in the criteria-selection phase, someone attempting to manipulate the outcome might emphasize the innate knowledge of a particular program required to effectively assess it, and offer up one of the only 'experts' available, possibly subject to personal bias.[15] And, just as narrowing the number of personnel considered qualified to conduct the

[13] For the range of possible biases and compromises evaluators face, including direct requests to favorably alter the results, see Kean, "Compromising Positions: The Objectivity of Evaluators," 87-88.
[14] Mislevy, "Validity by Design," 463-469.
[15] See note 12 above.

measurements and assessments, limiting the scope of an assessment or evaluation sample sizes would more easily allow a stakeholder to control the inputs and thus results.[16]

Further, the stakeholder could attempt to steer the assessment design of complex problems towards scripted events that could be rehearsed in advance and therefore not offer a true reflection of the program. Alternately, the stakeholder could advocate for an emphasis on self-reported or unverifiable information or contextual clues that might appear to give a qualitative indication of a program's status but really serve to disguise the true state.[17]

Stakeholders could also play to fiscal consciousness by raising the spectre of costs involved in assessing a program element in a particularly undesirable way, and conversely argue the thriftiness of those most desired. Likewise, the stakeholder could make the case that the costs of an overly thorough evaluation that brought in highly skilled experts with large sample sizes would be too high or unnecessary.

A final area of possible deception in a program stems from withholding, rather than mis-reporting. This is "the deliberate omission of relevant metrics, facts or issues related to the state of project activities,"[18] and can be used for the same goals of influencing perceptions or decisions related to a program. With this and the other avenues of deceit exposed, how can they be forestalled?

Combatting Deception and Lessons for Standards

[16] Mertens and Wilson, *Program Evaluation Theory and Practice.*
[17] Ibid.
[18] Smith, Thompson, and Iacovou, "The Impact of Ethical Climate on Project Status Misreporting," 577-591. Fulk and Mani, 1986. This study explored the "impact of organization ethical climate" on the likelihood of intentional misreporting, whatever the motive. As might be expected, the perception of an environment in which rules are followed strictly led to less misreporting by project members, while an environment in which project members are expected to act in an individually and self-interested way correlated with greater misreporting. The authors were somewhat surprised, however, to find that a "caring, team-spirited environment" had no discernible impact on misreporting probabilities.

While it is apparent that some stakeholders might have the desire and means to attempt to manipulate program assessments to hide the state of a program, those interested in accurate assessments can take steps to guard against these efforts. In large part, actions to combat deception through all stages of a program assessment simply involve remaining conscious of the aforementioned ways in which deception might occur and taking steps, such as the following, to counteract them when practical. Additionally, this same guidance need not be limited to instances where manipulation is expected, and can in fact strengthen the validity of any assessment to draw the most accurate picture of a program.

At the standards and criteria-establishment stage, wide stakeholder inclusion—frequently considered the key to achieving buy-in for a process— can help weed out invalid items in a form of peer review.[19] Determining how broadly to seek input can be tricky, but one option for those establishing standards or selecting criteria is to include several experts considered independent to evaluate and critique the possibilities. Additionally, industry standards for the characteristics of quality standards can serve as a guideline.[20]

In general, standards and criteria that should be given a priority include those that don't require overly expert knowledge to evaluate.[21] This would help prevent the opportunity for manipulation arising from personal bias. A possible consideration could be hiring or maintaining the skill set to competently perform the measurements or evaluations in an independent capacity, but such an approach could be prohibitively expensive.

Selecting standards that can be measured or ascertained directly, rather than second-hand through human communication, can also reduce the opportunities for deception. A 2002 study on detecting manipulation in IT systems determined that by not having to deal with a person who

[19] Schmidtz, "A Place for Cost-Benefit Analysis," 148-171.
[20] See for example Wholey, et al., *Handbook of Practical Program Evaluation*, 445.
[21] U.S. Government Accountability Office. GAO-12-208G.

might intentionally or unintentionally be deceptive ("the human factor") to receive data, evaluators may actually be more likely to detect deception—while at the same time generating a far lower level of "false positives."[22]

Lastly, the temptation for manipulating self-reporting can be counteracted by establishing standards in such a fashion that they facilitate verification.[23] This means choosing standards that themselves can be corroborated, such as easily quantifiable measurements, as well as those standards that can be verified through more than one data stream, preferably including an independent source. The GAO report on under-reporting income to Housing and Urban Development (HUD) was able to identify those engaged in deception through the use of third-party tax data reported to the IRS. Yet this approach is not without drawbacks, as the verification can be weighed down by cost and legal considerations, or as HUD found when the IRS pushed back against the pending policy, by bureaucratic prerogatives. As with most things it is a matter of tradeoffs, balancing more stringent standards guarding against deception with resource constraints.[24]

At the assessment-design level, stakeholder inclusion versus the independence of the assessors and process is another such tradeoff as previously discussed. The Methods Branch, one of the main branches of assessment methodology, recommends evaluators maintain distant relationships to help combat personal biases that may cloud their judgment.[25] Mertens and Wilson present evidence that this approach in conjunction with evaluation design sufficiently

[22] Biros, George, and Zmud, "Inducing Sensitivity to Deception in Order to Improve Decision Making Performance: A Field Study," 119-144. The term "false positive" describes a signal that something exists when it does not (a Type I Error in statistics), typically either an effect or a relationship. A false positive in deception detection would be a signal that a detection of deception has occurred when it has not in fact done so. Since false positives are errors, reducing them in this context prevents valid results from being discarded or discounted, enabling more accurate program assessments and program management.
[23] Martinelli and Parker, "Deception and Misreporting in a Social Program," 886-908.
[24] U.S. General Accounting Office. GAO/HRD-92-60.
[25] Mertens and Wilson, *Program Evaluation Theory and Practice*.

established to preclude such interpersonal effects can help negate attempts at manipulation, a view echoed by the U.S. GAO.[26]

As an example at the evaluation design stage, deception can be combatted by working towards the independence of evaluators. Over time Congress and the Executive have converged in thinking on the usefulness of this approach – from the implementation of inspector general programs insulated from political and managerial influence to the rise of independent cost estimates to the use of independent validation and verification of safety-critical DoD information technology systems during test and evaluation stages.[27] Using independent evaluators can be a multifaceted effort, such bringing in assessors from an outside agency to prevent an affective, fiscal, or factual compromise, and rotating assessors throughout the duration of the assessment so as to prevent an association compromise**.**[28]

Similar to the independent reviewers of proffered standards, it might be useful as costs permit to establish a "red cell." This small group would be tasked with identifying aspects of a project they would most want to hide or omit from reporting if they were trying to disguise a problem, thereby identifying additional criteria possibly useful for inclusion.[29]

Additionally, keeping the design process open to stakeholder input creates a further tension, with the program evaluation precept of the element of surprise. Ironically, one of the

[26] U.S. Government Accountability Office. GAO-12-208G.
[27] U.S. Department of Defense. Interim DoD Instruction 5000.2. MIL-STD-882E, "Standard Practice for System Safety," May 11, 2012
[28] Kean, "Compromising Positions: The Objectivity of Evaluators," 87-88. Cooley meanwhile argues that "complete objectivity" is neither possible nor desirable as it helps ensure the relevance of the assessment, but the specific nature of the evaluator's subjectivity must be transparent to all stakeholders and decision-makers (Cooley, "The Inevitable Subjectivity of Evaluators," 89-90). In practical effect the incorporation of evaluation compensations for an assessor's own averages and screening for conflicts of interest can help mitigate subjectivity.
[29] Smith, Thompson, and Iacovou, "The Impact of Ethical Climate on Project Status Misreporting," 577-591.

best ways to combat deception and gain a true reflection of a program may be deception in program design, in the form of surprises such as unannounced examinations.[30]

There are ethical considerations for the use of deception in any testing, including on the part of the evaluators to further the goals of the assessment.[31] Thayer and Padgett contend that "generally speaking, deception is not employed unless there is no other way to study the phenomenon, the phenomenon is scientifically important, and the risk of participation is minimal." If such criteria is applied to broader program evaluation it is entirely possible that deception on the part of the assessors for non-research programs can be both ethically viable and useful – especially if manipulation attempts are anticipated on the part of the assessed.

Assessor-on-assessed deception could entail surprise over the timing of the assessment, misdirection about what is being evaluated, or duplicity about the potential for the assessment to have a negative effect on the reputation or resources of those being evaluated. The common use of 'pop quizzes' suggests timing is a widely accepted practice. Going further and disguising not only the timing but also the intent may require more stringent controls, as reflected in the research field by the use of institutional review boards to prevent harm resulting from proposed deception. When the formality of a mechanism like an IRB is impractical, a possible compromise with stakeholder concurrence could entail agreement to include surprise elements within the program assessment, but leave undisclosed which aspects will be varied, such as the specific timing or program elements to be assessed.

As with standards establishment and criteria selection, a red cell could aid in identifying additional opportunities to combat deception by identifying tempting ways in which a

[30] Schmidtz, "A Place for Cost-Benefit Analysis," 148-171.
[31] Thayer and Padgett, *Program Evaluation: An Introduction*. Important elements of incorporating deception into evaluation the authors note are "consultation with others" – to verify the judgment that deception is necessary, useful, and will not harm – and that those deceived are subsequently fully debriefed, which could easily be folded into traditional program evaluation debriefs.

stakeholder could deceive. These might include guarding against a design that overly relies on contextual clues for indications of a program's effectiveness, confusing correlation and causation,[32] or a design that does not include a large enough sample size to avoid statistical or controlled anomalies.[33]

In Martinelli's low-income housing study, as the potential benefits in qualifying for the program increased, under-reporting increased and over-reporting decreased. This demonstrates that the likelihood of over- or under-reporting (whether in quantitative or qualitative[34] terms) can be modulated by linking the outcome of an assessment to an expected impact.[35] For example, a program evaluation designer worried about under-reporting of a negative effect could make it known that the those reporting the specific condition will in fact receive additional resources rather than solely be socially stigmatized—whether or not they actually will.

This is not approach is not without challenges. To accurately determine the modulation needed to balance under- and over-reporting it would be useful to run verifiable sample groups as in Martinelli's study. Further, modulation is not always available, whether because the impacts of the assessment are out of the assessment designer's control–or, if deception is to be used, so well known as to render deception unlikely to work.

When designing assessments to evaluate complex standards, Mislevy recommends observing several performances or multiple aspects of complex performances, including multiple observable variables. By increasing the amount of observations, assessors should have enough

[32] Believing in a cause-and-effect relationship (causation) when none exists due to the frequency of two variables occurring together (correlation). To return to our original example, while clean uniforms are frequently correlated to a ready military unity, because there is not typically a causal relationship between the two it can be misleading to rely on cleanliness as a sole means of determining readiness.
[33] Mertens and Wilson, *Program Evaluation Theory and Practice.*
[34] Craig, Mortensen, and Iyer examine the uses and promise of text analysis in identifying deception among program managers when given the ability to track changes among several instances of the same qualitative self-evaluations over time in "Exploring Top Management Language for Signals of Possible Deception," 333-347.
[35] Martinelli and Parker, "Deception and Misreporting in a Social Program," 886-908.

information to generate valid indications of the strength and nature of claims (e.g. that a program is sound, or that a program is effecting positive outcomes) when that data is fed through measurement models and probability distributions – for example that a program element would act in a desired way in a given situation, and that this outcome is the result of the positive efforts of the overall program.[36]

Once an assessment has been designed, one of the biggest aids for detecting deception is simply making evaluators aware of the possibility of deception. The authors of Martinelli's study further recommend training on detection deception as close as possible to the assessment to aid assessors' retention of their deception-detection abilities.[37]

Conclusion

From the stages of criteria and standards selection onward to evaluation design, program managers have a range of options to deceptively influence the outcome of assessments. Other stakeholders, and those wishing to mitigate and minimize manipulation, must remain on guard for its possibility and take proactive steps to reduce the possibility of deceit. These range from the use of open and transparent feedback to ensuring the independence of assessors to red cells identifying possible vulnerabilities.

As long as the stakes in a program assessment may influence decisions or influence perceptions, there is every reason to believe that some level of deception will continue in program reporting. Even when manipulation is unintentional, perhaps the result of unconscious prejudgment or preference, the effects on an assessment's outcome can be similar. Luckily, stakeholders interested in assessments as a true reflection of a program's state have a variety of

[36] Mislevy, "Validity by Design," 463-469.
[37] Martinelli and Parker, "Deception and Misreporting in a Social Program," 886-908.

methods at hand to mitigate their impacts. Even in assessments devoid of conscious deceit, the lessons drawn can help improve the fidelity and reliability of the evaluation's results. Yet as with much of the field, a lot of the recommendations are easier said than done.

Bibliography

Biros, David P, Joey F. George, and Robert W. Zmud. "Inducing Sensitivity to Deception in Order to Improve Decision Making Performance: A Field Study." *MIS Quarterly* (June 2002): vol. 26, no. 2: 119-144.

Cooley, William W. "The Inevitable Subjectivity of Evaluators." *Educational Evaluation and Policy Analysis,* (July-Aug, 1981): vol. 3, no. 4: 89-90.

Coryn, Chris L.S. and Carl D. Westine. "Eval 6000: Foundations of Evaluation." Western Michigan University. Powerpoint file dated September 23rd, 2010. Accessed December 17th, 2013. http://www.wmich.edu/evalphd/wp-content/uploads/.../September_23.pptx.

Craig, Russell, Tony Mortensen, and Shefali Iyer. "Exploring Top Management Language for Signals of Possible Deception: The Words of Satyam's Chair Ramalinga Raju." Journal of Business Ethics (March 2013): vol. 113, no. 2: 333-347.

Faulkner, William. "Critical System Heuristics." *Better Evaluation Beta*. Accessed September 15th, 2013. http://betterevaluation.org/plan/approach/critical_system_heuristics.

Habict, J.P., C.G. Victoria and J.P. Vaughn. "Evaluation Designs for Adequacy, Plausibility, and Probability of Public Health Programme Performance and Impact." *International Journal of Epidemiology* (1999): 28: 10-18.

Havens, Harry S. "Program Evaluation and Program Management." *Public Administration Review,* (July-Aug, 1981): vol. 41, no. 4: 480-485.

Kean, Michael H. "Compromising Positions: The Objectivity of Evaluators." *Educational Evaluation and Policy Analysis,* (July-Aug, 1981): vol. 3, no. 4: 87-88.

Martinelli, César and Susan Wendy Parker. "Deception and Misreporting in a Social Program." *Journal of the European Economic Association.* (June 2009): vol. 7, no. 4: 886-908.

MIL-STD-882E, "Standard Practice for System Safety," May 11, 2012

Mislevy, Robert J. "Validity by Design." *Educational Researcher*. (Nov 2007): vol. 26, no. 8: 463-469.

Mertens, Donna M. and Amy T. Wilson. *Program Evaluation Theory and Practice: A Comprehensive Guide*. New York: Guilford Press, 2012.

Thayer, Bruce A. and Deborah K. Padgett. *Program Evaluation: An Introduction*, 5th ed. Belmont, CA: Wadsworth, 2009.

Tolstoy, Leo. *War and Peace*. Translated by Richard Peaver and Larissa Volokhonsky. New York: Vintage Classics, 2008.

Patton, Michael Q. *Utilization-Focused Evaluation,* 4th ed. London: SAGE, 2008.

Schmidtz, David. "A Place for Cost-Benefit Analysis." *Philosophical Issues.* (2001): 11: 148-171.

Smith, H. Jeff, Ron Thompson and Charalambos Iacovou. "The Impact of Ethical Climate on Project Status Misreporting." *Journal of Business Ethics.* (2009): 90: 577-591.

U.S. Department of Defense. *Operation of the Defense Acquisition System.* November 25, 2013. Interim DoD Instruction 5000.02.

U.S. General Accounting Office. Report to the Chairman, Subcommittee on Housing and Urban Affairs, Committee on Banking, Housing, and Urban Affairs, U.S. Senate. Urban Poor. Tenant Income Misreporting Deprives Other Families of HUD-Subsidized Housing. GAO/HRD-92-60. Gaithersburg, M.D.: U.S. General Accounting Office, July, 1992.

U.S. General Accounting Office. Report to the Chairman, Subcommittee on Federal Services, Post Office and Civil Service, Committee on Governmental Affairs, US. Senate. Ballistic Missile Defense: Records Indicate Deception Program Did Not Affect 1984 Test Results. GAO/NSIAD-94-219. Washington, D.C.: U.S. General Accounting Office, July, 1994.

U.S. Government Accountability Office. *Designing Evaluations*. January 2012. GAO-12-208G.

Wholey, Joseph S., et al. *Handbook of Practical Program Evaluation, 3rd Edition*. San Francisco, CA: Jossey-Bass.